DOCUMENT RESUME

ED 327 582 TM 016 035

AUTHOR Chrzan, Keith
TITLE A Survey Version of Full-Profile Conjoint Analysis.
PUB DATE Aug 90
NOTE 15p.; Paper presented at the Annual Meeting of the American Psychological Association (98th, Boston, MA, August 10-14, 1990).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Adults; Analysis of Variance; *Data Collection; Estimation (Mathematics); Likert Scales; *Physicians; *Research Methodology; *Surveys
IDENTIFIERS Buyers; *Card Sort; *Conjoint Analysis; Rater Reliability

ABSTRACT

Two studies were conducted to test the viability of a survey version of full-profile conjoint analysis. Conjoint analysis describes a variety of analytic techniques for measuring subjects' "utilities," or preferences for the individual attributes or levels of attributes that constitute objects under study. The first study compared the developed survey to a card-sort approach for a large (27 profile) conjoint for which both ordinal and interval measures of the dependent variable were obtained. Subjects were 350 prospective buyers of a home medical instrument: 122 subjects received cards to sort and ranked the attributes, and 110 received the survey instrument and rated the profiles, first on a 5-point Likert scale and next on a 1-27 ordinal scale. (Return rates for these two groups were 61.5% and 73.3% respectively). Subjects in the second study were physicians attending a conference: 74 subjects completed a survey for 18 profiles, and 65 subjects completed a card sort to generate utility estimates. The method of data collection did not affect the conjoint utility estimates enough to alter the conclusions of these studies. Subjects' reliability affected utility estimates much more often than did the method of data collection. In both studies, the ordinal card sort procedure generated a greater proportion of reliable subjects than the survey method. In most circumstances, the card sort method seemed preferable. Four tables present the analysis of variance results. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED327582

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

KEITH CHRZAN

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

A Survey Version of Full-Profile Conjoint Analysis

by

Keith Chrzan

Presented at the American Psychological Association Convention, Boston, MA, August 13, 1990.

TM016035

Background

"Conjoint Analysis" describes a variety of analytic techniques for measuring subjects' "utilities," or preferences for the individual attributes or levels of attributes that constitute objects under study. If, for instance, a researcher wants to measure the relative importance in frozen yogurt brand selection of price, flavor and fat content, conjoint measurement would be appropriate. Flavor would be an attribute, chocolate a level and frozen yogurt the object.

Types of conjoint studies differ with respect to data collection method, utility estimation procedure and dependent variable measurement scale (Wittink and Cattin 1989, Cattin and Wittink 1982, Green and Srinivasan 1978). Common data collection methods include pairwise trade-offs between all possible level pairs or selected sets of attribute level pairs and trade-offs between "full-profiles," complete sets of attribute levels (Wittink and Cattin 1989). A full-profile of a frozen yogurt product would specify a level for each of the three attributes: $0.18/ounce, chocolate, 1 gram of fat/ounce.

Utilities can be estimated by a variety of programs, most notably MONANOVA (Kruskal 1965), LINMAP (Srinivasan and Shocker 1981) and ordinary least squares regression (OLS). Of these, MONANOVA has fallen from favor and OLS dominates the present (Wittink and Cattin 1989).

The dependent variables in a full-profile trade-off task are usually ordinal or interval scales (Wittink and Cattin 1989). Typically subjects sort decks of cards with a full profile on each card. If a subject ranks each card from most to least favorite, ordinal data results. If, on the other hand, a subject sorts cards into several Likert-type categories of preference, interval data may result. Researchers often have subjects sort by a two-stage process, first into Likert-type categories and then ranking within categories, so that both ordinal and interval data can result from a single card sort.

One of the limitations of full-profile data collection is the number of stimuli to which subjects must respond. Fractional factorial designs can greatly reduce the number of stimuli; they allow the

researcher to estimate main effects but few if any interactions (Addelman 1962, Green 1974, Green, Carroll and Carmone 1978). Green and Srinivasan (1978) suggest that subjects cannot sort more than about 30 full-profile cards, but some researchers suggest that as few as 12 full-profile cards may be the most subjects can handle (Carmone, Green and Jain 1978). Many of the problems to which a researcher may want to apply conjoint measurement involve numbers of factors and levels that require fractional factorial designs with more than even 30 stimuli. If the researcher wants to test for even a few interactions, the number of stimuli rises further.

To solve the problem of subject overload, the researcher may break the task into smaller portions and have separate groups of subjects sort separate portions. By means of "bridging" variables appearing in each smaller portion, the researcher can combine and possibly rescale the utilities from each portion after data collection (though the utilities of individual subjects are then lost). Another promising solution is "adaptive" conjoint analysis, though this requires the subject to interact with a computer terminal (Johnson 1987).

Seemingly the conjoint or trade-off aspect of the data collection method is valuable in forcing subjects to reveal their values by the trade-offs they make (Huber 1987). Not all of the data collection techniques, however, require trade-offs. If subjects need only sort stimulus cards into preference categories (possibly generating interval data) they need make no trade-offs at all, unless the distribution of their sort is constrained (e.g. of 18 cards and 5 categories, no more than 8 or less than 1 stimulus card per category).

One step further still from forcing trade-offs would involve turning the categorization task into a straightforward paper and pencil survey task. Profiles could be described several to a page and subjects could just rate them on a Likert type preference scale. A possible advantage of the survey conjoint is that it may be easier for subjects to rate larger numbers of profiles on a survey than they can sort in a deck of cards.

The ratio of observations to variables in conjoint studies tends to be small, so conjoint studies tend to operate with sparse degrees of freedom. When the ratio of observations to variables falls

too low and the residual degrees of freedom are too few, the coefficients of an OLS regression are unstable (Tabachnick and Fidell 1989, Hintze 1990, Mullet 1989).

Using the standard formula for significance of R^2 in a multiple regression, one can choose a significance level and find a critical R^2 for a given design:

$$R^2 = \frac{F_{\alpha,k,n-k-1}k}{(n-k-1) + F_{\alpha,k,n-k-1}k}$$

where n is the number of full profiles and k is the number of degrees of freedom of the independent variables. Subjects can then be classified as performing significantly better than chance or not; thus a new qualitative variable can be determined for each subject: subjects with significant R^2s can be considered "reliable" subjects and subjects with non-significant R^2s are "unreliable." Mullet (1989) notes that unmeasured interactions may also cause the relatively low R^2s, so that the measurement may be unreliable and not the respondents themselves). Unreliable subjects should be deleted from subsequent analysis because, in not supplying better-than-chance data, they add random error and bias if included (Chrzan, 1991). The ratio of reliable subjects to total subjects may be called the "good fit ratio" of a sample of subjects.

Current Studies

Two studies were conducted to test the viability of the survey version of full-profile conjoint analysis. OLS was chosen as the estimation program both because it is the most common method (Wittink and Cattin 1989) and because LINMAP and MONANOVA are more properly applied to ordinally scaled data (Green and Srinivasan 1978). The SPSS-PC version of OLS conjoint utility estimation was used on both studies.

The first study compares the survey conjoint to a card sort for a large (27 profile) conjoint for which both ordinal and interval measures of the dependent variable were obtained. The second study uses a smaller (18 profile) conjoint to compare the survey version of conjoint to a card sort providing only ordinal dependent measures. In both studies no assumptions were made about how

part-worth utilities for the levels of a given attribute were related to one another. A three level attribute thus became two dummy variables and absorbed two degrees of freedom. The hypotheses to be tested in both cases are that a) utilities derived from the survey do not differ from those derived from the more traditional card sort and b) that the two data collection techniques produce similar good fit ratios. Rejection of the former null hypothesis does not, however, imply the inferiority of the survey conjoint without further studies of the relative predictive validity of both methods.

Study I

Subjects

All subjects were prospective buyers of a durable home medical instrument drawn from a national consumer panel; 200 received card sorts and 150 received surveys. Useable returns numbered 122 (61.5%) and 110 (73.3%), respectively, but this difference owed more to the telephone response collection for the card sort as described below.

Instrument

The 27 profile fractional factorial design measured the main effects of a 9×3^2 full factorial experiment. Subjects sorting decks of full-profile cards sorted them first into 5 piles corresponding to 5 points on a Likert preference scale; next subjects rank-ordered cards within each pile for a 1-27 ordinal scale. Survey subjects simply rated the 27 profiles, presented in a survey format, on a 5-point Likert scale identical to the preference scale used in the first stage of the card sort. The profiles represented promotions of a class of durable home medical instruments. All subjects received their materials in the mail ; survey subjects returned their completed surveys *via* mail while card sort subjects were contacted by telephone for their responses, a feature of the study dictated by the organization financing the research.

Methodology

First, compute a critical value for R^2 such that subjects can be categorized as being reliable or not. Next, two-way ANOVA allows the effect of subject reliability to be separated from that of

method of data collection/dependent variable measurement scale. Finally, the good fit ratios of the card sort and survey samples are compared.

Results

Substituting n=27, k=12 and $F_{.05, 12, 14}$=2.53 into the equation for critical R^2 above showed that 95% of random cases would have R^2 of less than 0.684, so subjects with R^2 of less than 0.684 were classified as not performing significantly better than chance. The creation of this dichotomous reliability variable allows the effect of subject reliability to be separated from that of data collection method.

Three analyses were performed. First, survey (interval) derived utilities were compared to the interval card sort utilities using two-way ANOVA. Next, survey and ordinal card sort utilities were similarly compared and finally interval and ordinal card sort utilities were compared. In the latter cases utilities were standardized prior to comparison because the mean ranking was larger (13.5, necessarily) than the mean rating on a 5-point scale could possibly be.

Table I shows the results of the two-way ANOVA of interval utilities by reliability and data collection method. Two significant main effects for data collection method occurred among the 15 utilities, but subject reliability had significant main effects for eight of the 15 attribute levels. The eight interactions are also effects of reliability. Take part-worth 3 for example, with its significant ($p<0.001$) interaction and no significant main effects. Mean part-worths of 0 and -3 for the reliable card sort and survey subjects, respectively, are very close compared to the mean part-worths of their unreliable counterparts, -48 and 17 respectively. For both the card sort and the survey, unreliable subjects had a much different mean from reliable subjects, but in opposite directions: unreliable survey subjects had a much higher mean part-worth than did reliable survey subjects, while unreliable card sort subjects had a much lower mean than did the reliable card sort subjects. All eight significant interactions have this character. This is not surprising, however, because given that the unreliable subjects are *unreliable*, one would expect two groups of them to differ from reliable group means in opposite directions (thereby producing a significant interaction) as

often as in the same direction (thereby producing a significant main effect). So of the 18 significant effects, 16 owed to subject reliability and only two to method. Thus subject reliability more than data collection method produced differences in utilities for the survey *versus* interval card sort case.

The two samples did not produce significantly different good fit ratios.

Table II shows the results of the two-way ANOVAs of the 15 attribute levels for the survey *versus* ordinal card sort data. Utilities for three attribute levels showed significant main effects for method of data collection and 15 owed to subject reliability (four significant interactions between method of data collection and reliability of subjects and 11 main effects of reliability.

The two samples had significantly different ($Z=2.71$, $p<0.01$) good fit ratios: 84% of card sort subjects *versus* 69% of the survey subjects had better-than-chance fits.

Table III summarizes the results of the 15 two-way ANOVAs performed on the interval and ordinal card sort utilities. There were no attribute utilities with significant main effects for data collection method; there were five attribute levels with significant interactions of data collection method and reliability and seven for reliability, for a total of 11 significant effects due to subject reliability.

The good fit ratio for the card sort *qua* ordinal data was 84%, compared to 75% for the card sort *qua* interval data, a difference significant at $p<0.09$ ($Z=1.73$).

Discussion

While reliability and its interaction with data collection method still accounted for most of the significant differences (44), five significant main effects for data collection did appear. Clearly differences due to reliability vastly outnumber those due to data collection method.

Finally, the card sort data collection generated a greater proportion of reliable subjects than did the survey method, though this could owe to the greater amount of variation in the dependent variable available on a 27 point ranking scale than on a 5 point rating scale.

Unfortunately, the card sort subjects could change preference category in the course of rank ordering cards, so the reliability for the interval card sort subjects may be overstated. This design

flaw may be unimportant, however, because the card sort rankings generated higher good fit ratios than did the card sort ratings. If subjects are going to sort cards, they generate more reliable utility estimates from ranking the cards than from sorting them into Likert-type preference categories.

Study 2

Subjects

All subjects were physicians attending a medical convention. Useable complete data sets were obtained for 74 subjects for the survey and 65 subjects for the card sort.

Instrument

The 3^5 X 2 full factorial design was reduced to a main-effects only fractional factorial design requiring 18 full-profile stimuli. Because of the dominance of card sort rankings over card sort ratings in Study 1, card sort subjects only rank ordered cards - no card sort interval data was obtained. Profiles represented a class of possible future consumer durable medical instruments.

Methodology

Again the above equation established significance levels for R^2. Correlation analysis and two-way ANOVAs assess differences in utilities for the 16 attribute levels by data collection method and reliability as a dichotomous variable.

Results

With $F_{.05,11,6}$=4.03, n=18 and k=11, 95% of random cases have an R^2 of less than 0.881, so 0.881 was the cutoff separating reliable subjects from unreliable subjects.

Again the two sets of utilities were standardized to remove the scale effect and allow comparison *via* ANOVA.

As shown on Table IV, no significant effects due to method occurred, nor did any significant interactions of method and reliability appear. Three significant main effects of reliability did occur.

The good fit ratio of 58% for card sort subjects compared to 38% for survey subjects was significant at $p<0.02$ ($Z=2.42$).

Discussion

To an even greater extent than in the previous study, subject reliability, not method of data collection, influenced the conjoint utilities. The card sort again produced a greater proportion of reliable subjects, though again this may be due more to the choice of dependent variable scale than to the method of administration of the conjoint task.

Conclusions

First, method of data collection by itself does not affect conjoint utility estimates enough to alter the conclusions of most studies. Subjects' reliability affects utility estimates much more often. Method of data collection and reliability are related: the ordinal card sort data collection procedure generated a greater proportion of reliable subjects than did the survey methods in both studies. In most circumstances, then, the card sort task is preferable to the survey. When a) sample size is easy to acquire and when response rates are not important, or b) when the costs of card sorting are great, however, the survey does produce similar utility estimates if subjects' unreliability is controlled for by deleting unreliable subjects from the analysis.

Table I
ANOVA of Interval Utilities
by Method (card-sort, survey) and Subject Reliability*

	Mean Part-Worths						
	Reliable		Unreliable				
Part-Worth	Card Sort	Survey	Card Sort	Survey	p M x R	p Method	p Reliability
1	-94	-77	-50	-46	ns	.05	<.01
2	-44	-39	-83	-24	.05	ns	ns
3	0	-3	-48	17	<.01	ns	ns
4	103	95	24	69	<.01	ns	<.01
5	-18	-24	17	-27	.04	.05	ns
6	60	55	36	14	ns	ns	.001
7	43	39	83	21	.05	ns	ns
8	-50	[illegible]	20	-33	.01	ns	<.01
9	0	-2	1	8	ns	ns	ns
10	-39	-35	-24	-23	ns	ns	<.01
11	-9	-8	-11	-5	ns	ns	ns
12	48	42	35	28	ns	ns	.02
13	-20	-25	-24	-2	<.01	ns	.05
14	-5	-6	-2	-10	ns	ns	ns
15	25	31	26	11	.04	ns	.05
n	93	75	31	34			

*Two-Way ANOVA: utility by method of data collection and reliability.

Table II
ANOVA of Survey with Ordinal Card Sort Utilities by Method (card-sort, survey) and Subject Reliability*

	Mean Part-Worths						
	Reliable		Unreliable				
Part-Worth	Card Sort	Survey	Card Sort	Survey	p M x R	p Method	p Reliability
1	-146	-130	-52	-93	.02	ns	<.01
2	-81	-62	-41	-46	ns	ns	.02
3	-5	-5	-31	28	<.01	ns	ns
4	159	159	12	126	<.01	.02	<.01
5	-21	-48	-6	-57	ns	<.01	ns
6	112	91	29	27	ns	ns	<.01
7	81	65	66	50	ns	ns	ns
8	-89	-69	6	-60	<.01	ns	<.01
9	-9	0	16	26	ns	ns	.04
10	-66	-62	-28	-44	ns	ns	<.01
11	-14	-13	-9	-6	ns	ns	ns
12	80	75	38	50	ns	ns	<.01
13	-40	-43	-18	0	ns	ns	<.01
14	-8	-6	-9	-6	ns	<.01	<.01
15	46	55	24	25	ns	ns	<.01
n	104	75	20	34			

*Two-Way ANOVA: utility by method of data collection and reliability.

Table III

ANOVA of Card Sort Utilities by Interval Vs Ordinal and Subject Reliability*

	Mean Part-Worths						
	Reliable		Unreliable		p	p	p
Part-Worth	Interval	Ordinal	Interval	Ordinal	M x R	Method	Reliability
1	-136	-146	-113	-52	<.01	ns	<.01
2	-64	-81	-59	-41	ns	ns	ns.
3	-4	-5	-33	-31	ns	ns	<.01
4	151	159	98	12	<.01	ns	<.01
5	-26	-21	-15	-6	ns	ns	ns
6	94	112	78	29	.03	ns	<.01
7	61	81	80	66	ns	ns	ns
8	-73	-89	-47	-6	<.01	ns	<.01
9	-4	-9	1	16	ns	ns	ns
10	-60	-66	-55	-28	ns	ns	.04
11	-16	-14	-3	-9	ns	ns	ns
12	75	80	58	38	ns	ns	<.01
13	-29	-40	-49	-18	.02	ns	ns
14	-8	-6	-4	-6	ns	ns	ns
15	38	46	54	24	.03	ns	ns
n	93	104	29	20			

*Two-Way ANOVA: utility by level of data and reliability.

Table IV
ANOVA of Utilities, Interval Vs Ordinal by Method (card-sort, survey) and Subject Reliability*

	Mean Part-Worths				p	p	p	
	Reliable		Unreliable					
Part-Worth	Card Sort	Survey	Card Sort	Survey		M x R	Method	Reliability
1	102	96	61	74		ns	ns	ns
2	-7	21	20	8		ns	ns	ns
3	-95	-117	-80	-82		ns	ns	ns
4	-15	-18	-40	-28		ns	ns	ns
5	0	18	43	23		ns	ns	.04
6	15	0	-3	4		ns	ns	ns
7	-9	-29	-21	-24		ns	ns	ns
8	-7	9	5	20		ns	ns	ns
9	16	20	16	4		ns	ns	ns
10	-45	-35	-61	-72		ns	ns	.05
11	14	0	14	-1		ns	ns	ns
12	31	35	46	73		ns	ns	.04
13	-59	-49	-40	-35		ns	ns	ns
14	59	49	40	35		ns	ns	ns
15	80	74	72	60		ns	ns	ns
16	-28	-20	-36	16		ns	ns	ns
17	-52	-54	-36	-76		ns	ns	ns
n	38	28	27	46				

*Two-Way ANOVA: utility by method of data collection and reliability.

References

Addelman, S. (1962) "Orthogonal Main-Effect Plans for Asymmetrical Factorial Experiments," Technometrics, **4** (February), 21-46.

Carmone, F., Green, P. and A. Jain (1978) "Robustness of Conjoint Analysis: Some Monte Carlo Results," Journal of Marketing Research **15**, 300-303.

Cattin, P. and Wittink, D. (1982) "Commercial Use of Conjoint Analysis: A Survey," Journal of Marketing **46** (Summer), 44-53.

Chrzan, K. (1991) "Unreliable Respondents in Conjoint Analysis: Their Impact and Identification," Sawtooth Software Conference Proceedings, Forthcoming.

Green, P. (1974) "On the Design of Choice Experiments Involving Multifactor Alternatives," Journal of Consumer Research, **1** (September), 61-68.

Green, P., Carroll, J. and Carmone, F. (1978) "Some New Types of Fractional Factorial Designs for Marketing Experiments," Research in Marketing, **1**, 99-122.

Green, P. and Srinivasan, V. (1978) "Conjoint Analysis in Consumer Research: Issues and Outlook," Journal of Consumer Research, **5** (September), 103-123.

Hintze, W. (1990) "The Degree of Freedom Problem in Conjoint Analysis," Proceedings of the Advanced Research Topics Forum, forthcoming from the American Marketing Association.

Huber, J. (1987) "Conjoint Analysis: How We Got Here and Where We Are," Sawtooth Software Conference Proceedings, 237-251..

Johnson, R. (1987) "Adaptive Conjoint Analysis," 1987 Sawtooth Software Conference Proceedings, 253-265.

Kruskal, J. (1965) "Analysis of Factorial Experiments by Estimating Monotone Transformations of the Data," Journal of the Royal Statistical Society, Series B, **27**, 251-263.

Mullet, G. (1989) "On Conjoint Studies with Scarce Degrees of Freedom: Is there Enough Utility to Go Around?" Marketing Research, **1** (June) 24-30.

Srinivasan, V. and Shocker, A. (1973) "Linear Programming Techniques for Multidimensional Analysis of Preferences, Psychometrika 38 (September) 337-69.

Tabachnick, B. and L. Fidell (1989) Using Multivariate Statistics 3rd Edition, (New York: Harper and Row).

Wittink, D. and Cattin, P. (1989) "Commercial Use of Conjoint Analysis: An Update," Journal of Marketing, **53** (July), 91-96.